DOCUMENT RESUME

ED 247 318 TM 840 540

AUTHOR Thorndike, Robert L.
TITLE Adaptive Testing: More Bang for the Buck.
PUB DATE 83
NOTE 26p.; Paper presented at the Annual Meeting of the Rocky Mountain Educational Research Association (Tucson, AZ, November 2-5, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Viewpoints (120)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Adaptive Testing; Computer Assisted Testing; Difficulty Level; Latent Trait Theory; *Student Evaluation; Test Items; *Test Use
IDENTIFIERS Two Stage Testing

ABSTRACT

In educational testing, one is concerned to get as much information as possible about a given examinee from each minute of testing time. Maximum information is obtained when the difficulty of each test exercise matches the estimated ability level of the examinee. The goal of adaptive testing is to accomplish this. Adaptive patterns are reviewed for three types of testing situations: (1) group testing, (2) interactive testing with a computer, and (3) individual testing by a trained examiner. Special attention is given to the stradaptive test format. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED247318

ADAPTIVE TESTING:

MORE BANG FOR THE BUCK

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
[ ] This document has been reproduced as received from the person or organization originating it.
[X] Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
R. L. Thorndike
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

by

Robert L. Thorndike

Paper presented at the 1983 annual meeting of the Rocky Mountain Educational Research Association.

TM 820 510

Abstract

In educational testing, one is concerned to get as much information as possible about a given examinee from each minute of testing time. Maximum information is obtained when the difficulty of each test exercise matches the estimated ability level of the examinee. The goal of adaptive testing is to accomplish this. Adaptive patterns are reviewed for three types of testing situations: (1) group testing, (2) interactive testing with a computer, and (3) individual testing by a trained examiner. Special attention is given to the stradaptive test format.

Adaptive Testing: More Bang for the Buck

Robert L. Thorndike

I have given my remarks the subtitle "more bang for the buck". This may need a little interpreting. In the present context "bang" is to read "information" and "buck" is decoded as "minute"-- more information per minute of testing. For the purpose of testing is to obtain the information that will help us with the decisions that we need to make, and we can always use more and better information than we have time to obtain.

Perhaps Willie Witless and Susie Smart entered our school yesterday from an Armed Services dependent school when their mothers were moved back to the States, and we must decide in what grade they should be placed and which section of that grade each belongs in -- whether the regular section , a special group of the intellectually gifted, or a program for slow learners. We feel that information about their general academic aptitude will help us decide. And in addition, we might reasonably like to know how well each understands what they read, and what competence each has with arithmetical concepts. Perhaps Ho Chi Min has just arrived in our school from Saigon by way of Rangoon, and we want to find out what command he has of English as well as how good he is at reasoning with quantitative and with non-verbal concepts. Perhaps Freddie Featherbrain and Connie Confused have shown up at the Guidance Office asking what they should plan on doing when they complete high school, so we would like to get a reading on verbal, quantitative, spatial, mechanical and clerical abilities for each of them. Time is often limited, and there are usually a number of things we need to know about a person if we are to make sound decisions. Furthermore, we need information that is accurate as possible, so that we will not be led astray by a

faulty reading. So that procedures that enable us to obtain more information per minute of testing time are eminently to be desired. .

When we speak of getting information, of course we mean NEW information, adding to our available supply of information about the person we are studying. No new information is gained if a test exercise merely provides the answer to a question to which we already knew the answer. Thus, if Sally was a high school senior, we would not ask her "How much is 2 plus 2?" When she responded "Four" we wuld have learned nothing new about her. We already knew, with a probability of about 999 out of 1000 that a high school senior would be able to give the correct answer to that question. Of course, if she had answered "Five, dummy.What do you think I am? Stupid?" we would have found out that she was a wiseacre and that she didn't think too much of the silly question we were asking -- but we wouldn't really have added to our information about her level of arithmetical ability. Conversely, if Willie was a second grader and we had asked him "What is the cube root of 27?" and he had answered "Ah-duh", we would have learned nothing new about Willie. Without any testing, we could have affirmed, at well above the .99 confidence level, that a second grader wouldn t know anything about cube roots. (Of course, if Willie had promptly said "Three", we would appear to have gained a great deal of information,to wit, that this second grader appeared to be one in a million. And we should promptly undertake to confirm this impression by asking "What is the cube root of sixty four?") -- But the probability of getting this response is SO small as to make asking the question a

generally unprofitable expenditure of testing time.

The point is that we gain information from a person's response to a test item to the extent that we were uncertain beforehand what the response would be. And the amount of information that we gain is in proportion to our initial uncertainty. If we feel initially that there is a fifty-fifty chance that Willie can answer a particular question, that constitutes maximum uncertainty. Then the result that he does (or does not) give the proper answer gives us a maximum amount of new information about him. The ideal test for Willie would be one where each question was one for which we thought beforehand that there was about a 50-50 chance that he could get it right.

In aptitude testing, and to a considerable extent also in achievement and personality testing, we think of a test as assessing some underlying trait, or latent attribute of the individual. We are inclined to give the attributes labels such as verbal ability, mechanical knowledge, reading comprehension, and so forth to communicate what we have in mind. The point is that we think in terms of a relatively unified aspect of persons. And we envisage a continuous scale for that attribute along which both test tasks and individual examinees can be placed. The goal of a test is to determine as quickly and as accurately as possible where on that scale a person falls.

Before any test is given we always have SOME notion of where along the ability dimension an individual falls. We might only know that Willie is a ten-year-old, and then the best we could do

would be to place him where an average ten-year-old would fall. We would place him there with a good deal of tentativeness, because ten-year-olds are quite a varied crew, but this estimate would still be a good bit better than a COMPLETELY random guess. We might know in addition that his former teacher had recommended that Willie be given some tests because he was having difficulty in keeping up with his class and was doing poorly in academic work. And on that basis we might decide to place him somewhere in the bottom quarter of ten-year-olds, and to feel somewhat increased confidence in our estimate because of the additional information that we had, though still recognizing that there remained a pretty wide band of uncertainty. We might know further that on a group ability test given in the third grade Willie had fallen at the rifth percentile on the norms provided by the test publisher, and we might use this fact to further modify and sharpen our initial estimate of where Willie stands now.

Whatever its nature and amount, we always have SOME previous information about a child that gives us SOME initial basis for estimating his level on the attribute that we are studying. This PRIOR estimate should never be considered an exact and certain placement at a specific point. It should really be thought of as the mid-point of some range within which we feel confident that his ability falls -- essentially a frequency distribution with a mean and a variance. We speak of this as a prior probability distribution of the status of that person-- our estimate prior to getting further information from a test. The purpose of testing is to produce a POSTERIOR probability distribution -- one that takes account of the new evidence -- in which the mean is

adjusted as appropriate on the basis of the test performance , and in which the variance is reduced as rapidly as possible , that is, in which the precision of our estimate is increased as rapidly as possible.

It is not only the ability of the individual that falls somewhere on the scale of the latent attribute. For Willie we speak of his ABILITY level. For a task such as "What is 1/2 divided by 1/4?" we speak of its DIFFICULTY level, and this difficulty level CAN also be located along the same scale of the attribute. This relation of a test exercise to the scale of the attribute is evidenced by what is called the "item characteristic curve". The item characteristic curve shows the probability that an examinee will get that item right as a function of where the examinee's ability falls on the scale of the attribute.

Item characteritic curves for several test items are illustrated on the chart in the hand-out.The item characteristic curve is ordinarily represented by the normal ojive, or by its very close relative, the logistic function. Both of these functions start at near zero percent for examinees with very low levels of ability, increase at first slowly as ability goes up and then more rapidly, and flatten out as they approach 100 percent at high ability levels. The item characteristic function has two parameters. One, its steepness, corresponds to the effectiveness with which the item differentiates between persons at lower and higher levels of the attribute. In the hand-out curves A and C are steep while B, D, and E are relatively flat. The other parameter is the location of the item's 50 percent

point on the scale of the attribute, a parameter that serves to define the difficulty level of the item. Here A and B are relatively easy items, while C and D are difficult.

We get the greatest amount of new information about Willie from an item that has a steep item characteristic curve, and one whose 50 percent point falls AS NEAR AS POSSIBLE to his estimated ability level . Such an item will be passed by a relatively large proportion of persons falling above the level of our prior estimate and a relatively small proportion of persons falling below that point. So if our prior estimate of Willie's ability was that he fell at about -2.0 on the ability scale in the figure, and if Willie passed item A, there would be a strong probability that he really falls above rather than below -2.0, and we should then adjust our prior estimate upwards. Thus, we can try at each step in testing to estimate the level at which Willie falls and to give him test items whose difficulty matches that level.

Of course, unless we have a good deal of prior information about Willie, our initial estimate of his ability level is usually very rough, and may be considerably in error. We shall need to modify it, perhaps substantially, on the basis of test results as we go along. Thus, using our best a priori guess, we choose and administer a test item. If Willie passes the item, we must raise our estimate of his ability level a bit, and then administer a somewhat more difficult item. We continue giving more and more difficult items until Willie fails an item. With each item passed, we must not only set a higher value as our

estimate of Willie's ability level, but we may have more confidence in the accuracy of that estimate.

After a few increasingly difficult items, Willie will ordinarily fail an item. At that point, we must lower our posterior estimate a bit and so we will drop back and give a slightly easier item. Soon we will have bracketed Willie's ability level, and will be giving items that are all close to that level, where Willie will be maitaining about an even balance of passes and failures. We will be getting a maximum amount of information from each item, and narrowing down the band within which we can be confident that Willie's ability lies.

The scenario is ssentially the same if Willie fails the initial items. We then drop down progressively, until we come to an item that is passed, and as before, we soon reach a level at which successes and failures occur about equally often. We are once again bracketing the level at which Willie functions, and our posterior estimate locates him at this level with progressively greater precision.

As with any testing enterprise, we must have a pool of items whose difficulty level is known, having been previously established in the course of test development. Given such a pool of items, we can design a strategy of item selection that will zero in for each examinee on those items whose difficulty will be appropriate for him, and will provide the maximum amount of information about him. Experience indicates that with the optimum choice of items, a specified degree of precision in our

evaluation of a person can be achieved giving only about half the number of items that would be needed if they were presented in a conventional test.

To present the ideal adaptive test we must have access to a computer that permits us to do the following. Firstly, the complete pool of useful items must be stored in the computer's memory. Secondly, the computer must be programmed so that for a given prior estimate of the level of the examinee's ability it will search the item pool, locate the item that provides the maximum amount of information at that level, and present it on the screen to the examinee. Thirdly, it must be programmed to record and score the examinee's response, and on the basis of that response to generate an appropriately revised posterior estimate of the examinee's ability level and of the precision of that estimate. Fourthly, it must be programmed to repeat this cycle of searching for the best item, administering that item, and updating the estimate of the examinee's ability level until some criterion for terminating testing is reached. This criterion may be some specified level of precision for the estimate, some specified number of items administered, or exhaustion of the supply of appropriate items. Finally, it must stop and report the final posterior estimate, usually with a statement of its precision.

Initially, the adjustment that we need to make in our estimate of Willie's ability level after each item is fairly substantial, since we started out with minimal information about Willie and each item passed or failed represented an appreciable

increment to the total of our information. However, as more items have been given, each new item represents a smaller and smaller proportional increment to our total supply of information. So adjustments in our estimate become progressively smaller, and oscillations in that estimate eventually can become small indeed.

Adjustments in our estimate arise both from items passed and items failed. I have spent several years trying, with somewhat limited success, to convince a colleague that one does get information about an individual's ability as much from his failures as from his successes. I still don't think he is really persuaded. He still seems to feelthat information about an examinee is only provided by that examinee's successes. As a matter of fact, with multiple choice items one gets MORE information from failures than from successes. A right answer may represent nothing but a lucky guesss, but a failure is a failure is a failure.

This computer-administered routine works out fine if we have the computer resources that permit us to put into the computer's memory the item pool and the program for selecting the items and presenting them to the examinee, and if we can then put the examinee in front of a terminal for interactive administration of the test. I had always thought of this as representing quite a chunk of computer. But as a matter of fact, it doesn't take that elaborate a computer installation to achieve a fairly close approximation to the ideal that I have just described. As I was organizing my ideas for this talk, I had a weekend visit from my

son-in-law, who cut his baby teeth on an Apple and he ran off a program for my micro that will store a test composed of as many as 15 items at each of as many as 15 difficulty levels, and another program that will adjust the difficulty level of the next item presented to an examinee each time he passes or fails an item. An Apple or its equivalent isalready available in ell over half of the schools in our country, and it may soon be as common as a typewriter, so this would be a widely useable program.

Let me illustrate how this program would work with a mini-test that I put together just to try it out. The items for this test are shown on the front of the second page of the handout. In this little test there are only five levels of difficulty, and five items at each level. The program permits the examinee to start at whatever level he or she chooses -- or more likely in an actual test, at the level that the examiner directs. The computer screen displays the first item at that level, and the examinee punches the number corresponding to his or her choice of answer. If Sally chooses the correct option, the computer then displays tne first question from the next HIGHER level, while if she choosees a wrong answer the computer displays the first item from the next LOWER level. Each time she gives a correct answer, the computer moves up a level, while each wrong answewr causes the computer to drop down a level. The first unused item in a colrmn is always the one to be given. If Sally reaches the top level and gets the item right, she just stays there, gets the next item at that level, and continues on at that level.

The test continues until all the items in some column have been used up. At this point, the testing stops, and the computer records the number of items from each level that were administered, the number that Sally got right, and the average difficulty level of those that she got right. This average could constitute the raw score.The score would need to be referred to some group norms for interpretation, as would be the case with any conventional test.

Of course, to be useful for any practical testing a test would need to have more levels and more items in each level than in this illustration, but these would exist as items in the pool in the memory of the computer. Only a small fraction of the total would need to be administered to any one examinee. You may find it interesting to try this tiny test on a spouse, colleague or enemy, to see how many of the 25 items actually get used in arriving at that persons score.My impression is that you will find that not more than half of the items are used, these being predominantly from 2 or 3 of the 5 levels. If the test were expanded to 15 levels, so as to cover the complete range of vocabulary difficulty, with 15 items at each level, the typical examinee would probably be tested with only about 40 of the pool of 225 items, and would be located within quite narrow limits on that full range of word knowledge.

The type of test that I have just described has been called by Dave Weiss of the University of Minnesota, who has done a great deal of work on adaptive testing, a stradaptive test, because of its adaptive nature and its arrangement of items in

strata according to difficulty level. This is a relatively simple adaptive format, and one that can fit into even the most limited computer system. The complete Bayesian testing design that I described earlier is probably a bit more efficient. However, the stradaptive format comes very close to the optimum in efficiency. It is a simple design and one that can be widely applicable in any sort of small computer installation.

But we can't always count on having a computer at hand not all kinds of items can be scored by a computer, and we can't always test examinees one at a time. What can we do if no computer is available, or if we must test all the pupils in a school at once? How can we make group testing adaptive, or what can we do to adapt to the single examinee when the adaptation must be carried out by a human examiner with all the limitations of judgment, attention, memory, and patience that afflict the typical examiner?

The simplest way to adapt test difficulty to pupil ability is to produce a test in a multi-level format. In this format a test or a subtest is composed of successive modules each composed of perhaps 4 or 5 items . with each module containing more difficult items than the preceding one. One level of the test for administration is composed of 4 or 5 consecutive modules, and different levels of the test are defined by different starting and stopping points.

Thus, in our Cognitve Abilities Test there are 12 modules for each subtest, and these provide 8 levels of the test, each composed of 5 consecutive modules. For a subtest, Level A might be composed of items 1 to 25, Level B of items 6 to 30, and so

on. A particular level is designed for the typical child at a certain grade level. The goal was to choose the items so that perhaps 90 percent of a typical group would get each item in the initial module correct, 60 percent would get each item in the middle module, and perhaps 30 percent in the final module. But of course, a higher or lower level may be administered when prior information indicates that the group -- or the individual pupil -- falls above or below what is typical for that grade. For complete classes, out-of-level testing is very simple -- but it is only moderately effective because of the range of talent that still remains in any classroom. For the single pupil, "out of level" testing is administratively more bothersome because children must be moved around to get those who are to take a particular level in the same room, but this can permit closer adaptation to each child.

At best, a multi-level format is somewhat limited in the amount of adaptation that is possible, because each test level must still cover a rather wide range of difficulty in order to be effective for the whole range of abilities found in a group testing situation. One cannot design the test to focus on just the ability level that would be most efficient for a particular child. But out-of-level testing does permit at least a gross adjustment of test tasks to the group's ability. What more might we do?

One possibility that has been explored somewhat is to use a basic multi-level format, but to let the examinee determine which levels he will take. The procedure would be likely to operate by

starting all examinees at a level that would be moderately difficult for the typical member of the group being tested. Willie and Susie would each complete the initial module, and each would then decide whether to go up to a more difficult module or down to an easier one. This would continue until a specified number of modules had been taken, with the only constraint being that they be contiguous modules. In the mini-test that I gave you, we might start everyone with Level 3, and specify that three levels be taken. Willie Witless could then take levels 3, 2, and 1; Susie Smart levels 3, 4, and 5; and Mary Middleton levels 3, 2 and 4.

The self-selection procedure places a fairly heavy demand on the examinee to follow directions, and to comprehend that the final evaluation of his or her performance will depend as much upon the level of tasks attempted as upon the preportion of successes. It is natural to think that you will get a better rating if you can do all the items, and not easy to communicate the notion tht you may make out better in the end if you push yourself to your limit and even beyond. Where this procedure has been tried out with college students, it appeared that they tended to elect items that were too easy to be most informative about them. They were inclined to play it safe.

Another approach to adaptive testing is that of multi-stage testing. The simplest version of this approach is two-stage testing. I can illustrate this approach by once again using a short vocabulary test. On the back of the hand-out you will see a 20-item vocabulary test. This is a steeply graded test

composed of two words from each of ten levels of the Intelligence Scale CAVD. I persuaded George Gallup some 40 years ago to include it in one of his surveys of American adults. The average score was about 11 right, while 3 percent got scores of zero and half a percent got scores of 20. It really spread this group out. But the point is that if we had an adult coming in off the street for counseling or as an applicant for employment, we might not have much idea what level of test to give him. If we wanted to give the Cognitive Abilities Test, should it be very easy, perhaps Level D to adjust to his limited education and/or ability? Should it be of moderate difficulty, perhaps Level F? Or is this a smart cookie who can only be adequately assessed by a stiff, demanding test such as Level H? Even this 5-minute 20-word mini-test that you have before you has a reliability of about .80 in a general adult sample and would provide a guide that would enable us to proceed with further testing with the assurance that it would be reasonably appropriate for that examinee.

One minor problem with two-stage testing is that the initial routing test must be scored before we can determine what level of further test should be given. This is no real problem when a single individual is being tested, because the short routing test can be scored in a few seconds, but it does present a little difficulty in group testing. For large scale testing programs, one solution is to make use of the latent image answer sheets that have been developed in recent years. These are based on a chemically printed answer sheet and a chemically treated felt-tipped pen. When the two chemicals interact, the image

printed on the paper becomes visible. If the correct answer space is coded "C", then when the test has been completed, it is only necessary to count the number of "C's" appearing in the spaces that were marked. Each examinee can score his own test. Based on that count, each examinee can be directed to the level of the main test that he is to work on.

The idea of two-stage testing isn't really new. Examiners have been applying a version of it for years in the administration of the old stand-by Stanford-Binet. It has been rather usual practice to start that test, at least for school age children, by giving the Vocabulary test. The examinee's level of performance on that sub-test, together with chronological age, guides the examiner to the levels of the Binet that will bracket the examinee's ability level. Two-stage testing is a simple and reasonably effective way to increase our prior information, so that we can avoid wasting time on test exercises that are so easy as to be trivial or so difficult as to be frustrating for the examinee.

For individual testing in which the test is being administered by a trained examiner, a promising refinement of two-stage testing can be to combine the use of an initial routing test with the modular multi-level format that we condidered earlier. In this procedure, the routing test will be used to determine the entry level into the modular test. If the examinee passes more than half of the items in the initial module, the examiner will move up and administer the next more difficult module and will continue up until the specified number

of modules has been administered. If half or less of the items in the first module are passed, the examiner will proceed downward to easier modules. The final scale score assigned to an individual will depend upon the levels administered combined with the number of items passed. This will guarantee a reasonable balance of uccesses and failures for each examinee. This procedure permits making a separate adjustment for each subtest involving a different type of test task, and permits an ability profile to be obtained with a maxiumum of precision and a minimum

There is one further advantage of an adaptive test format that Iwould like to point out at this time. This is that every examinee, no matter at what ability level, will have roughly the same proportion of successes as any other examinee. The items may be at a lower level, but the least able examinee will get to experience nearly the same proportion of successes as the most able. To the extent that the adaptive pattern has been well designed, both Willie Witless and Sally Smart will be passing about half the items presented to them.

The implications of this fact for examinee motivation are profound. Studies exploring adaptive formats have found that examinees like them -- and well they might. The slow child, in particular, is no longer faced with repated failure and frustration. And the very bright child is quickly brought up to tasks that challenge his or her ability. For examinee motivation and morale, this feature of the adaptive test is perhaps even more important than its greater efficiency in providing information about a child.

There are a number of other models that have been proposed for adaptive testing. One of these arranges items in a pyramid, in which difficulty spreads out as one goes down through the layers of the pyramid. An examinee starts at the peak, and works down taking one item from each layer, his route being determined by the pass or fail on the previous layer. Another model is an alternation format in which the items are arranged in a V, one branch gradually increasing in difficulty, and the other progressively decreasing. The examinee starts at the point of the V and goes up the upper limb until he fails an item. At that point he switches to the lower limb and goes down it until he passes an item. At this point he shifts to the first unused item in the upper limb. This alternation continues with each failure and pass. It is my impression that neither of the above patterns are as effective as the ones that I described earlier.

The best choice from among the several models depends upon whether testing is in groups or individual, and whether all the items are ones that have completely objective scoring.

If the testing must be done in groups, the multi-level modular test seems to be the best we can do. With mature and fairly sophisticated examinees, this might be combined with a two-stage procedure that would permit the examinee to determine on an individual basis which levels he should take.

If scoring is objective, so that a computer can handle it, if testing can be individual, and if a micro-computer instalation is available, the stradaptive test format would be practical and would be a close approximation to the ideal.

When the test exercises are ones the scoring of which requires an examiner's judgment, then a two-stage or multi-stage format would seem to be the best. It permits a quick and flexible focussing of testing at about the ability level of the examinee with a minimum of lost time and a minimum of frustration and failure.

In the field of international relations it has been asserted that the buck has too much bang these days, and that our international trade is being ruined by the strength of the dollar. But in the field of testing, the trade-off is all in favor of more bang -- that is, information -- for the buck -- that is, minute of testing time.

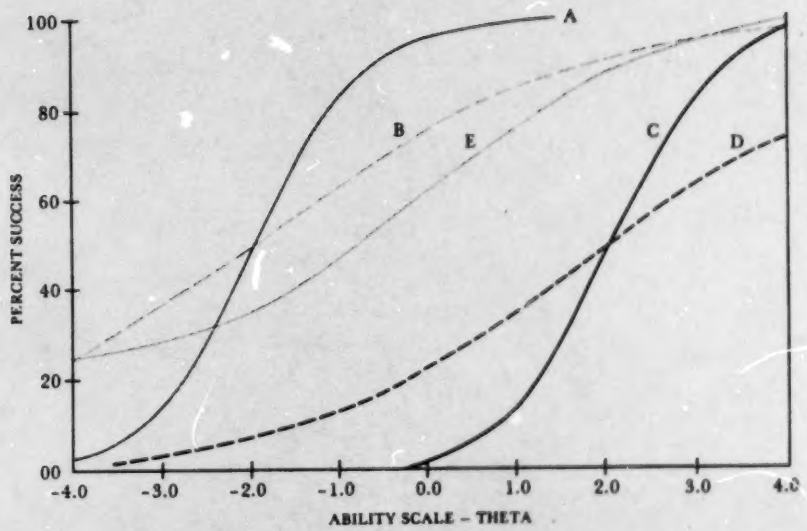

Figure 1.2 Selected item characteristic curves.

SWC 1

Stradaptive Word Knowledge Test

LEVEL 1	Level 2	LEVEL 3	LEVEL 4	LEVEL 5
1. PROLOGUE	1. GRATUITY	1. FALLACIOUS	1. REPOSE	1. UNTOWARD
1-conclusion	1-friendliness	1-obscene	1-rephrase	1-away from
2-conversation	2-elegance	2-misleading	2-rest	2-in new direction
3-sales talk	3-tip	3-repetetive	3-ask again	3-hostile
4-speaker	4-thankfulness	4-insecure	4-change position	4-unfortunate
5-introduction	5-politeness	5-lively	5-reproduce	5-mistaken
2. PARCH	2. DOCILE	2. COVERT	2. DIVEST	2. TENET
1-sit	1-stupid	1-lonely	1-entertain	1-basic belief
2-speak out	2-tame	2-obvious	2-produce	2-resident
3-dry	3-can be bent	3-secret	3-remove	3-joint
4-piazza	4-tenth	4-persuade	4-redirect	4-grip
5-regret	5-free	5-antagonistic	5-vary	5-legal proceding
3. FAMISHED	3. INCANDESCENT	3. ALLEGE	3. RETROACTIVE	3. TESTATOR
1-disappeared	1-continuous	1-claim	1-acting back on	1-giver of tests
2-tired out	2-inexpensive	2-deceive	2-radioactive	2-maker of tests
3-polished	3-electric	3-border	3-acting slowly	3-maker of a will
4-replaced	4-glowing	4-break down	4-contradictory	4-measuring device
5-starved	5-flickering	5-sensitivity	5-very sensitive	5-container
4. DIVERT	4. CREDIBLE	4. SODDEN	4. DOGMATIC	4. MILKSOP
1-distract	1-breakable	1-immediate	1-self centered	1-sissy
2-regret	2-believable	2-drunk	2-unfriendly	2-luncheon dish
3-produce	3-impossible	3-soaked	3-canine	3-disease of cattle
4-disagree	4-honest	4-hairy	4-pig headed	4-baby food
5-discharge	5-good to eat	5-grass covered	5-religious	5-lazy person
5. TERMINAL	5. INCISION	5. DOMICILE	5. PHILANTHROPY	5. CHARLATAN
1-lasting	1-uncertainty	1-camel	1-charitable giving	1-loose woman
2-final	2-conclusion	2-friendly	2-study of culture	2-clown
3-abrupt	3-section	3-tame	3-therapy	3-tropical bird
4-distant	4-cut	4-servant	4-solitariness	4-fraud
5-resting place	5-prediction	5-home	5-stinginess	5-dancer

Key: 1-1 (5); 1-2 (3); 1-3 (5); 1-4 (1); 1-5 (2); 2-1 (3); 2-2 (2); 2-3 (4); 2-4 (2); 2-5 (4)
3-1 (2); 3-2 (3); 3-3 (1); 3-4 (3); 3-5 (5); 4-1 (2); 4-2 (3); 4-3 (1); 4-4 (4); 4-5 (1)
5-1 (4); 5-2 (1); 5-3 (3); 5-4 (1); 5-5 (4)

SIDE 2

Vocabulary Routing Test

Directions: For each word in capital letters, circle the word below it that means most nearly the same thing.

1. SPACE
   1. school
   2. noon
   3. captain
   4. room
   5. board

2. LIFT
   1. sort out
   2. raise
   3. value
   4. enjoy
   5. fancy

3. BLUNT
   1. dull
   2. drowsy
   3. deaf
   4. doubtful
   5. ugly

4. NATION
   1. business
   2. daughter
   3. country
   4. chief
   5. animal

5. ADVERTISE
   1. detain
   2. explore
   3. give notice of
   4. adverse
   5. newspaper

6. BOG
   1. ebb
   2. disorder
   3. swamp
   4. field
   5. difficult

7. EDIBLE
   1. auspicious
   2. eligible
   3. fit to eat
   4. sagacious
   5. able to speak

8. PACT
   1. puissance
   2. remonstrance
   3. agreement
   4. akillet
   5. pressure

9. CONSPIRE
   1. plot
   2. breathe
   3. rely
   4. die
   5. outrun

10. CONCUR
    1. agree
    2. race
    3. mongrel
    4. pounce
    5. ramble

11. ALLUSION
    1. aria
    2. illusion
    3. eulogy
    4. dream
    5. reference

12. CAPRICE
    1. value
    2. a star
    3. grimace
    4. whim
    5. inducement

13. MADRIGAL
    1. song
    2. mountebank
    3. lunatic
    4. ribald
    5. sycophant

14. AVARICE
    1. ordinary
    2. verious
    3. empress
    4. frailty
    5. greed

15. CONFLAGRATION
    1. carnival
    2. celebration
    3. fire
    4. contagion
    5. decorate with flags

16. RECIPROCAL
    1. saturnine
    2. mutual
    3. receptive
    4. morose
    5. careless

17. PRISTINE
    1. flashing
    2. earlier
    3. primeval
    4. bound
    5. green

18. TACTILITY
    1. tangibility
    2. grace
    3. subtlety
    4. extensibility
    5. manageableness

19. LARGESS
    1. enormity
    2. present
    3. monstrosity
    4. amiability
    5. size

20. MONOMANIA
    1. flying machine
    2. conceit
    3. one-colored
    4. endogen
    5. aberration

KEY: 1.(4), 2.(2), 3.(1), 4.(3), 5.(3), 6.(3), 7.(3), 8.(3), 9.(1), 10.(1), 11. (5), 12.(4), 13.(1), 14.(5), 15.(3), 16.(2), 17.(3), 18.(1), 19.(2) 20.(5)